# The pyramid quantized Weisfeiler-Lehman graph representation

Katerina Gkirtzou
Research Center "Athena", Marousi, Greece

Matthew B. Blaschko
KU Leuven, Leuven, Belgium

Women in Machine Learning 2015

**Abstract**

Graphs are flexible and powerful representations for non-vectorial data with inherited structure. Exploiting these data requires the ability to efficiently represent and compare graphs. Unfortunately, standard solutions to these problems are either NP-hard, hard to parametrize or not expressive enough. Graph kernels, that have been introduced in the machine learning community the last decade, is a promising solution to the aforementioned problems. Graph kernels have been shown to enable efficient and accurate statistical learning, but many graph kernels have high order polynomial time complexity. Furthermore, efficient graph kernels rely on a discrete node labeling as a central assumption. However, many real world domains are naturally described by continuous or high-dimensional vector valued node labels. Despite the significant progress in the design and improvement of graph kernels, efficient and expressive representation and comparison of graphs with such complex labels remains an open research problem.

We help to address this problem by proposing a novel framework, the *pyramid quantized Weisfeiler-Lehman graph representation* [1], an efficient graph representation and comparison scheme for large graphs with continuous vector labels. Our algorithm considers statistics of subtree patterns with discrete labels based on the Weisfeiler-Lehman algorithm. The key advantage of these subtree statistics, which are tree structures constructed recursively from each node in the graph up to a predefined depth $h$, is their linear complexity to the number of edges in the graph. Moreover, they make use of an efficient hashing scheme enumerating the relevant dimensions of an exponentially sized feature space. To take advantage of this efficient scheme when working on continuous vector labeled graphs, we use a pyramid quantization strategy to determine a logarithmic number of discrete labelings that results in a representation that guarantees a multiplicative error bound on an approximation to the optimal partial matching. As a result, we approximate a graph representation with continuous vector labels as a sequence of graphs with increasingly granular discrete labels. We evaluate our proposed algorithm on two different tasks with real datasets, on a fMRI analysis task and on a generic problem of

3D mesh shape classification. In the fMRI analysis task, we approach the problem by representing the fMRI contrast maps as graphs with continuous labels. Our primary hypothesis is that the interconnections between voxels can contain additional information about the brain structure, that could not be explored under a linear consideration of the data. Our proposed method validates this hypothesis and outperforms other machine learning techniques with statistical significance. In the 3D shape classification task, we view the 3D shape models as graphs enriching their nodes with labels from local features, such as curvature of the surface. Our method performs better that the pyramid Bag of Words approach, showing that the spatial structure is important for improving the classification performance. Moreover, the results show that our approach contains complementary information to the multi-viewpoint rendering technique, since the combination of both approaches performs better with statistical significance than each method individually. Source code of the implementation can be downloaded from `https://web.imis.athena-innovation.gr/~kgkirtzou/Projects/WLpyramid.html`.

# References

[1] Katerina Gkirtzou and Matthew B. Blaschko. The pyramid quantized Weisfeiler-Lehman graph representation. *Neurocomputing*, 2015.